DOCUMENT RESUME

ED 070 659 24 SE 015 468

AUTHOR Harris, Margaret L.; Romberg, Thomas A.
TITLE Measuring Mathematics Concept Attainment: Boys and Girls.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
REPORT NO WURDCCL-TR-195
BUREAU NO BR-5-0216
PUB DATE Nov 71
CONTRACT OEC-5-10-154
NOTE 34p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Ability; Division; *Elementary School Mathematics; *Evaluation; *Mathematical Concepts; Mathematics Education; Relationship; *Research; Set Theory; Student Evaluation; Test Construction; Testing; Tests

ABSTRACT

Test development efforts for constructing 12 items to measure achievement of each of 30 selected mathematics concepts are described. The topics covered are sets, division and expressing relationships. Item and total score statistics for data collected on 196 girls who had just completed the fifth grade during early summer of 1970 and 195 boys who had just begun the sixth grade during the fall of 1970 are presented and discussed. For related documents, see SE 015 462 and SE 015 469. (Author/DT)

ED 070659

TECHNICAL REPORT NO. 195

MEASURING MATHEMATICS CONCEPT ATTAINMENT: BOYS AND GIRLS

REPORT FROM THE PROJECT ON A STRUCTURE OF CONCEPT ATTAINMENT ABILITIES

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FILMED FROM BEST AVAILABLE COPY

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR COGNITIVE LEARNING

THE UNIVERSITY OF WISCONSIN
Madison, Wisconsin

U. S. Office of Education
Center No. C-03
Contract OE 5-10-154

SE 015 468

ED 070659

Technical Report No. 195

MEASURING MATHEMATICS CONCEPT ATTAINMENT: BOYS AND GIRLS

By Margaret L. Harris and Thomas A. Romberg

Report from the Project on
A Structure of Concept Attainment Abilities
Robert E. Davidson, Lester S. Golub, Herbert J. Klausmeier,
Thomas A. Romberg, B. Robert Tabachnick, Alan M. Voelker
Principal Investigators

and
The Quality Verification Program
Mary R. Quilling, Director

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

November 1971

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

Statement of Focus

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Quality Verification Program and from the Project on the Structure of Concept Attainment Abilities in Program 1. The Quality Verification Program assisted in developing tests to measure concept achievement and identifying reference tests for cognitive abilities, while the Concept Attainment staff took primary initiative in identifying basic concepts in mathematics at intermediate grade level. The tests will be used to study the relationships among cognitive abilities and learned concepts in various subject matter areas. The outcome of the Project will be a formulation of a model of structure of abilities in concept attainment in a number of subjects, including social studies, science, and language arts, as well as mathematics.

Contents

		Page
	List of Tables and Figures	vii
	Abstract	ix
I	Introduction	1
II	Procedures	5
	Test Development	5
	Readability	5
	Validity	5
	Reliability	6
	Item Revision	6
	Subjects	8
	Data Collection	8
	Treatment of the Data	9
III	Results and Discussion	11
	Reliability Estimates and Test Statistics	11
	Item Indices	13
IV	Summary and Conclusions	27
	Conclusions	27
	Recommendation	27
	References	29

List of Tables

Table		Page
1	Mathematics Concepts Categorized by Area	3
2	Test Data for the Population and Samples of Madison, Wisconsin Fifth Grades	9
3	Distribution of Fathers' Occupations	10
4	Means, Standard Deviations, and Reliabilities for Mathematics Concept and Task Scores: Boys and Girls	12
5	Reliability Estimates for Task Scores by Area and Total for Girls	13
6	Item Indices Based on Concept and Task Criterion Scores	14

List of Figures

Figure		
1	Item matrix for each individual.	7

Abstract

Test development efforts for constructing 12 items to measure achievement of each of 30 selected mathematics concepts are described. Item and total score statistics for data collected on 196 girls who had just completed the fifth grade during early summer of 1970 and 195 boys who had just begun the sixth grade during the fall of 1970 are presented and discussed.

I
Introduction

The primary objective of the project entitled "A Structure of Concept Attainment Abilities" (hereafter referred to as the CAA Project) is to formulate one or more models or structures of concept attainment abilities, and to assess their consistency with actual data. The major steps for attaining this primary objective were taken to be:

1. To identify basic concepts in language arts, mathematics, science, and social studies appropriate at the fourth grade level,

2. To develop tests to measure achievement of these concepts,

3. To identify reference tests for cognitive abilities, and

4. To study the relationships among learned concepts in these four subject matter fields and the identified cognitive abilities.

This paper describes the test development efforts for measuring achievement of selected concepts in mathematics; thus, it is a report of one aspect of Step 2. As such, it will include descriptive item and test statistics for the tests developed. The items can be found in "Items to Test Level of Attainment of Mathematics Concepts by Intermediate-Grade Children" (Romberg & Steitz, in press).

Concepts may be defined in one or more of four ways: (a) structurally, in terms of perceptible or readily specifiable properties or attributes; (b) semantically, in terms of synonyms or antonyms; (c) operationally, in terms of the procedures employed to distinguish the concept from other concepts; or (d) axiomatically, in terms of logical or numerical relationships (Klausmeier, Harris, Davis, Schwenn, & Frayer, 1968). "A concept exists whenever two or more distinguishable objects or events have been grouped or classified together and set apart from objects on the basis of some common feature or property of each" (Bourne, 1966, p. 1). The concept of Bourne's definition might be called a classificatory one and seems to be the same as the structural type discussed by Klausmeier, et al. (1968). This is the type of concept with which this project is concerned, and such a definition of a concept served as the basis for selection and analysis of subject matter concepts.

Many different types of performance might be taken as the critical evidence that a student does or does not understand a given concept. Thus, as a part of this project it is necessary to have a schema for measuring understanding of concepts. Such a schema was developed by Frayer, Fredrick, and Klausmeier (1969) and was used by the CAA Project to assess concept attainment. The "Schema for Testing the Level of Concept Mastery" consists of 13 types of questions, each involving a different task required of the examinee. The schema also allows for selection of an answer (multiple-choice type questions) or for production of an answer (completion type questions). It was decided to use the first 12 tasks and a multiple-choice format for this project. The 12 tasks of the schema which were used are:

1. Given the name of an attribute, select an example of the attribute.

2. Given an example of an attribute, select the name of the attribute.

3. Given the name of a concept, select an example of the concept.

4. Given the name of a concept, select a nonexample of the concept.

5. Given an example of a concept, select the name of the concept.

6. Given the name of a concept, select the relevant attribute.

7. Given the name of a concept, select the irrelevant attribute.

8. Given the definition of a concept, select the name of the concept.

9. Given the name of a concept, select the definition of the concept.

10. Given the name of a concept, select the supraordinate concept.

11. Given the name of a concept, select the subordinate concept.

12. Given the names of two concepts, select the relationship between them.

Single- or compound-word classificatory concepts (those that are defined by attributes) in mathematics subject matter at the fourth grade level were identified. This task was subdivided into four steps:

1. Identification of the major areas within the subject matter of mathematics,

2. Selection of three of these major areas to be studied,

3. Identification of classificatory concepts within each of these three major areas, and

4. Random sampling of ten concepts from those identified for each of the three major selected areas.

This yielded a total of 30 mathematics concepts to be studied by the project. A list is given in Table 1, by area, of the concepts identified. The areas are Sets, Division, and Expressing Relationships. In a pilot study, it was found that a very small percentage of mid-year fourth grade students could pronounce or render any meaning to nine of the concepts in the area of Division. They are algorithm, associative property, closure property, commutative property, density property, distributive property, identity property, order property, and reciprocal property. These concepts were excluded from the random sampling procedure. A description of the procedures used to identify these concepts can be found in "Selection and Analysis of Mathematics Concepts for Inclusion in Tests of Concept Attainment" (Romberg, Steitz & Frayer, in press). The researchers of Project 101, Situational Variables and Efficiency of Concept Learning, developed a system for analyzing a concept in preparation for developing items to measure the level of attainment of that concept (Frayer, Fredrick, & Klausmeier, 1969). Since the publication of that paper they, in cooperation with the researchers of the CAA Project, have refined their thinking and advanced this system. The refinements are discussed in "A Structure of Concept Attainment Abilities: The Problem and Strategies for Attacking It" (Harris, Harris, Frayer, & Quilling, in press). Briefly, a concept may be described in many ways--in terms of its criterial, relevant, and irrelevant attributes; its examples and nonexamples; its supraordinate, coordinate, and subordinate hierarchical relationships (theoretically determined); and its lawful or other types of relationships to other concepts. Knowledge of each of these kinds of information may be tested to determine a student's level of attainment of a concept. An analysis, along these lines, of each of the 30 sampled mathematics concepts which are being studied can be found in "Selection and Analysis of Mathematics Concepts for Inclusion in Tests of Concept Attainment" (Romberg, Steitz & Frayer, in press).

Thus, using the analysis of a concept as the basis for appropriate content and the 12 tasks of the schema as the basis for appropriate tasks, 12 items, one for each of the 12 tasks, were developed wherever possible for each of the 30 concepts. For seven of the concepts, no item was developed for Task 11, so there was actually a total of 353 rather than 360 mathematics items for the purpose of measuring and assessing concept attainment in mathematics. The development of the items, along with item and total score statistics (for concepts and for tasks) obtained for them for fifth grade boys and girls, will be discussed in the following sections.

Table 1
Mathematics Concepts Categorized by Area

Sets	Division	Expressing Relationships
Cardinal Number	Algorithm	Area
*Disjoint Sets	Associative Property	*Average
Element	Closure Property	Dozenal System
*Empty Sets	Common Denominator	Estimation
*Equal Sets	Commutative Property	Generating Sentences
*Equivalent Sets	*Denominator	*Graph
Intersection	Density Property	Length
*Line	Distributive Property	Liquid
Line Segment	*Division	Mathematical Sentences
Non-Disjoint Sets	*Factor	*Measurement
Ordered Pairs	*Fraction	*Open Sentence
*Parallel Lines	Identity Property	Partial Sums
*Plane	*Mixed Fraction	*Place Holder
*Point	*Multiplication	*Place Value
Set	*Numerator	Range
Sets of Numbers	Order Property	Round Numbers
Sets of Points	Partial Product	*Solution Set
Skew	Partial Quotient	*Standard Unit
*Subset	Partitioning	*Statement
*Subtraction - A way of looking at addition	*Product	*Weight
Triangular Numbers	*Quotient	
Union of Sets	Reciprocal Property	
Universal Set	*Remainder	
Whole Number		

* Concepts randomly selected to be tested.

II
Procedures

This section contains a discussion of the item development procedures used including initial item construction and revision of those items based on item analysis results. Also included is a discussion of the data collection procedures, subjects, and treatment of the data.

Test Development

One item for each of the 12 tasks was generated for each of the 30 selected concepts. If one looks at the tasks being used to measure understanding of the concept, it is apparent that there can be more than one item generated for at least some of the tasks. For example, a Task 1 type item could be constructed to measure understanding of each of many relevant attributes for most concepts. For this project, it was decided to construct just one multiple-choice item for each task for each concept. This made it necessary to have bases for making choices when such choices were necessary. These bases consisted of principles for selecting attributes, relationships, incorrect choices, etc. A discussion of such bases may be found in "A Structure of Concept Attainment Abilities: The Problem and Strategies for Attacking It" (Harris et al., in press).

General procedures for item construction included initial item generation by a subject matter specialist item writer; critique of the items by a committee composed of the item writers from each of the four subject matters being studied (the other three are language arts, science, and social studies), an experienced elementary school teacher specializing in reading, and a measurement specialist; and final critique by the subject matter principal investigator and a measurement specialist. Concerns in the item construction process were readability, validity, and reliability.

Readability

It was intended that no student should be unable to answer an item correctly simply because of inability to read the item. In writing items, very simple language was used wherever possible. Several pilot studies concerned with the readability question were conducted, and two outside consultants expert in the testing and measurement fields were asked to look at a sample of the items from the point of view of readability for fifth graders. No significant differences were found among treatment groups; percentage of occurrences of subjects who could not pronounce the word and did not know its meaning when shown the concept labels, but did know its meaning when the word was pronounced, was judged to be negligible; and the two outside consultants independently advised that there was no reading problem with the items and that there should be no concern about administering them in the standard way in which the students read the items themselves. The conclusion drawn from the results of the pilot studies and the consultants' opinions was that readability of the items was not a problem and under standard administration conditions would be satisfactory. For further information see Harris et al. (in press).

Validity

The content validity of each of the items was of immediate concern during item construction; aspects of construct validity were to be probed later using duplicate test construction, simplex analysis, and factor analysis of the results obtained using the content-valid items

constructed.

Content Validity. Each item was constructed to meet the content and task specifications set for it. The task required of the student by each item was specified by the schema adopted for use in measuring concept attainment. The concept name was given by the sampling process; the attributes, examples, definition, and relationships associated with the concept name were defined by the prior analysis of the concept. The content for each item was specified in this manner. The content specifications were not as precise as the task specifications due to the necessity of choosing a single attribute to be tested for example and selecting the incorrect alternatives to be used in the multiple-choice questions. Systematic construction of alternate choices was used whenever possible; for example, for an item dealing with the operation of addition, the operations (or examples of them) of subtraction, multiplication, and division were used as incorrect choices.

To further ensure the content validity of the items, two persons who were familiar with the schema for testing concept attainment, but were not involved in the item development process, classified five random sets of 72 items (12 items for six concepts in each set) according to content and task. These two persons had the analyses of the concepts available. They were able to correctly classify all but a few of the items. Any questions they had about these few items were mutually resolved among the subject matter principal investigator, the measurement specialist, and themselves.

Reliability

Developing one item for each of the 12 tasks for each of the 30 selected concepts yields a 12 (tasks) by 30 (concepts) matrix consisting of the score for each of the 360 items, one for each cell of the matrix, for each individual to whom the items were administered. Thus, a completely crossed design exists and two types of total scores can be secured from this matrix: a total score for each of the 30 concepts (totalled across tasks) and a total score for each of the 12 tasks (totalled across concepts). Figure 1 is an illustration of such a matrix.

This design offers these alternatives: (a) use a total score of 360 items to analyze all items against; (b) use 30 total scores, each for one concept and consisting of 12 items, to analyze the 12 task items against; and (c) use 12 total scores, each for one task and consisting of 30 items, to analyze the 30 concept items against. The first alternative was rejected since it assumes neither task nor concept variation is present. A choice was not made between the next two alternatives. Instead, both were done. An important theoretical problem of how to item analyze a completely crossed design like this remains to be solved.

Major concerns about reliability for the test development process were that internal consistency reliability estimates for task scores (total of 30 items across concepts) and concept scores (total of 12 items across tasks) be high enough to warrant further study using such scores. It was recognized that there might be some contradictions in what was attempted. The items were constructed to comply with the completely crossed design, 30 concepts by 12 tasks. One major objective of the entire project is to determine the dimensionality of the selected mathematics concepts and of the tasks when using mathematics content. If either or both of these are not unidimensional, then an internal consistency reliability estimate based upon items measuring aspects from the multidimensions would reflect this; the more dimensions present and the more uncorrelated they are, the lower the internal consistency estimate. Recognizing this, and not being able to study the dimensionality of the two modes (concepts and tasks) until after the items were developed, pilot studies were conducted using the items for some of the concepts for the 12 tasks. As will be pointed out later, evidence indicates that sufficiently reliable scores can be obtained for both task scores and concept scores.

Item Revision

If one looks at the 12 tasks for a single concept it becomes quite apparent that there may be a strong learning effect as one attempts to answer the items. The name of the concept appears in every item, except for the first two which deal with an attribute of the concept, either in the stem or as a possible choice. This makes a random presentation of the items desirable. Using items for six of the mathematics concepts presented on mark sense type cards, a study was conducted in which one group of subjects responded to the items arranged in the same random order (over 72 items for the six concepts) common to all subjects. The second group of subjects

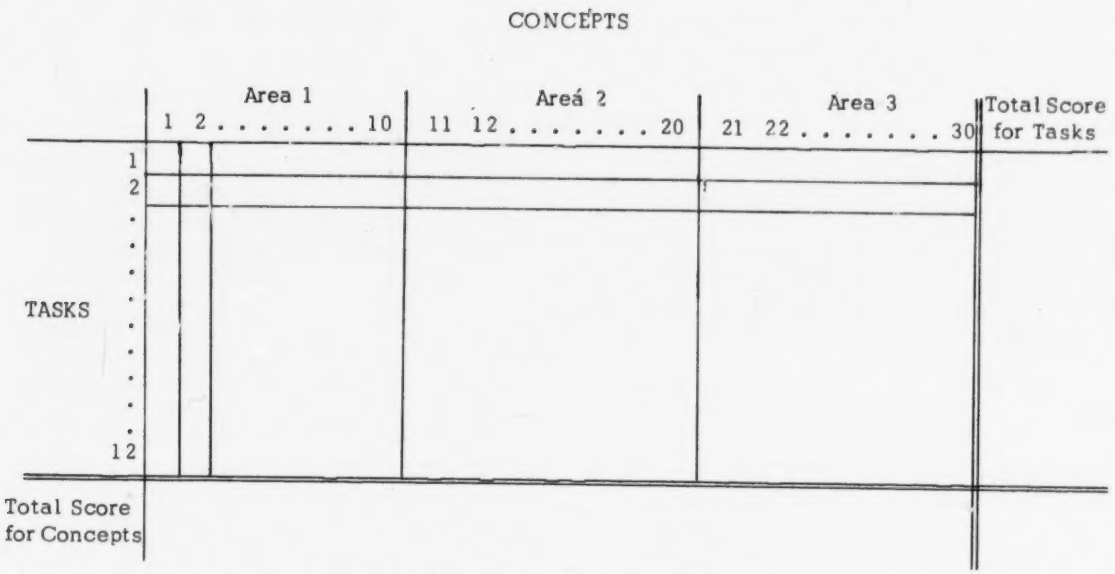

Fig. 1. Item matrix for each individual.

responded to the items arranged in a random order (over 72 items for the six concepts) which was a unique one for each subject of the group. No significant differences in test score were found between the subjects receiving a common random order and those receiving a unique random order.

Tryouts of the items for item analysis and revision purposes were conducted using a single random order over the items for six concepts contained in a test booklet. This constituted a "test" of 72 items which could readily be administered in 1 hour. The tryouts were conducted during December, 1969, and January, 1970, with fifth grade students in the Madison, West Allis, and Fond du Lac, Wisconsin school systems. All of these school systems used the Greater Cleveland Mathematics Program. Approximately 100 students responded to each "test." Madison students responded to the items for six of the concepts, West Allis the items for 12, and Fond du Lac students the items for 12 of the concepts.

The tryout data were subjected to the Generalized Item Analysis Program (GITAP) (Baker, 1969), the output of which provides the proportion responding, item-criterion biserial correlation, X_{50} (point on the criterion scale corresponding to the median of the item characteristic curve), and β (the reciprocal of the standard deviation of the item characteristic curve which is a measure of the discriminating power of the item) for each possible choice for each item as well as summary descriptive statistics for the total test. It also gives the Hoyt reliability for the total test and the standard error of measurement.

As discussed earlier, the design for these mathematics achievement items is one in which the concepts and tasks are completely crossed. Since there are no item analysis procedures available for completely crossed designs, the data were analyzed in each of the two possible ways--each item as part of the appropriate concept score and as part of the appropriate task score. This raises questions as to the interpretation of such results. The main referents used for interpreting the results and as a basis for making item revisions were the results obtained from the analyses of the concept scores. The tasks were fixed and thus any arbitrary decisions were made in regard to appropriate content for incorrect choices, etc. Usual standards for item indices were not strictly adhered to, as a unique design for item analysis was being used and a major objective of the project is to study the dimensionality of the concepts and of the tasks. If high discrimination indices were demanded, the dimensionality might have been affected by making the items more homogeneous. Also, no attempt was made to manipulate the difficulty level of the items, since another objective of the project is to determine if any differential levels of difficulty, or complexity, exist in the concepts

and in the tasks. Therefore, the item analysis results were used as a very general guide to help in determining whether there were "hidden" weaknesses, clues, and/or incongruities in the items and, in an even more general sense, to show that what we were attempting to do was possible--sufficiently reliable concept and task scores could be obtained when using this completely crossed design.

The revised items can be found in "Items to Test Level of Attainment of Mathematics Concepts by Intermediate-Grade Children" (Romberg & Steitz, in press).

Subjects

The mathematics items were administered to 196 girls who had just completed the fifth grade during early summer, 1970, and to 195 boys who were just beginning the sixth grade during the fall of 1970 in the public school system of Madison, Wisconsin. The students were randomly selected from the population of all such girls and from the population of all such boys. The Madison Public School System made available the information concerning the populations and used their computing facilities to designate the random sample for the girls.

Initially, a random sample of 300 girls was drawn. Letters were sent to the parents of these students explaining the purpose and details of the testing, and inviting their daughter to participate in the testing program. A stamped and addressed postcard was enclosed which the parents were asked to complete and return indicating whether or not they were willing to allow their daughter to participate. One hundred and two yes responses and 25 no responses were obtained from the cards returned. Those parents who had not returned the card by a specified date were phoned. An additional 46 yes and 61 no responses were obtained by phone. Since this total of yes responses did not give as many subjects as were desired, an additional sample of 150 girls was drawn at random. From this sample, 56 yes and 30 no responses were obtained by card. Thus, of the total sample of 450 students, 203 yes and 116 no responses were received; seven students did not complete the testing, which resulted in a total of 196 girls tested. These students were paid $7.50 for participating.

A random sample of 756 boys was drawn and letters were sent. By mail, 420 yes and 87 no responses were obtained. Thirty-eight of the subjects did not complete the testing, resulting in 382 boys tested. Of this total, 195 boys completed the mathematics and social studies items; the others responded to language arts and science items. As with the girls, the boys who completed the testing program were paid $7.50.

Since the participation of all students comprising the random sample was impossible to attain, test and IQ data were obtained from the files of the Madison Public School System for both the school population and those participating students for whom the information was available. Table 2 includes the summary statistics for the population of fifth grade students in the public school system of the city of Madison during the school year of 1969-70, and for the boys and the girls who comprised the tested samples for the mathematics items. The IQ scores were obtained in a fall, 1968, administration of the Lorge-Thorndike Intelligence Test when the subjects were fourth graders; and the scores on the Iowa Tests of Basic Skills, given in grade equivalent scores, were obtained in the fall of 1969 when the subjects were fifth graders.

Data on fathers' occupations were collected from the students using the Master Occupational Code of the United States Bureau of the Census. These data were tabulated and are presented in Table 3.

Data Collection

The data for the girls were collected in two different schools during five 2-hour daily sessions for one week. Subjects could choose the week and the school in which they wanted to report for testing. A one-week session was held at Hawthorne School from June 22 to June 26, and a one-week session was held at Hoyt School from July 13 to July 17. Each 2-hour session consisted of a 72-item "test" composed of mathematics items, a 72-item "test" composed of social studies items, and an activity break between the two of approximately 1/2 hour. The mathematics and the social studies items were given first on alternate days.

The data for the boys were collected in a similar manner from mid-October to mid-November. Ninety of the boys who were attending Middle School for sixth grade were tested after school for five consecutive days in one week at Schenk, Sennett, and Orchard Ridge schools; those 105 elementary school boys who completed the testing (who were attending a Junior High School) were tested on three consecutive Saturday mornings at Franklin, Longfellow, and Randall schools.

Table 2
Test Data for the Population and Samples
of Madison, Wis. Fifth Grades

		Population	Boys	Girls
Lorge-Thorndike Intelligence Test	$\bar{X}$	106.60	105.95	112.02
	s		14.74	12.15
	N	2605	169	191
Iowa Tests of Basic Skills				
Vocabulary	$\bar{X}$	5.53	5.60	5.75
	s		1.39	1.34
	N	2520	181	187
Reading Comprehension	$\bar{X}$	5.44	5.43	5.84
	s		1.60	1.46
	N	2520	181	187
Language Skills	$\bar{X}$	5.24	5.07	5.74
	s		1.43	1.29
	N	2520	181	187
Work-Study Skills	$\bar{X}$	5.46	5.50	5.70
	s		1.31	1.13
	N	2520	181	187
Arithmetic Skills	$\bar{X}$	5.05	5.08	5.24
	s		1.04	.97
	N	2520	179	187
Composite	$\bar{X}$	5.35	5.34	5.65
	s		1.22	1.10
	N	2520	179	185

The mathematics items were arranged in five 72 item "tests." The order of the items was assigned randomly over the 360 items. Two different random orders were used to collect the data: one for each school for the girls and one for each type of school for the boys.

The items were arranged in five test booklets according to the random order. The students responded to the items by marking their chosen response directly on an answer sheet. The answer sheets were read by machine and the responses punched onto data cards.

Treatment of the Data

The treatment of the data consisted of two main procedures: reliability estimation and item analysis. The data were analyzed separately for each sex group. Hoyt analysis of variance reliability estimates were obtained for each of the 30 concept scores and each of the 12 task scores for each group studied. Means and standard deviations for each of the scores were also computed.

Item analyses using the GITAP program (Baker, 1969) were obtained for each of the items as a part of two different scores: an appropriate concept score and an appropriate task score. This program provides proportion responding, item-criterion biserial correlation, X_{50}, and β statistics for each choice of each item. The proportion of students who respond correctly to an item is an index of the difficulty level of that item. The greater the value of the difficulty index, the easier the item. The biserial correlation coefficient is an index of the discriminating ability of the item choice. For these analyses the criterion ability used was total concept or total task score. X_{50} is the point on the criterion scale, given in standard deviation units, corresponding to the median of the item characteristic curve. It is the point at which subjects with that score have a 50-50 chance of choosing that response. β is the reciprocal of the standard deviation of the item characteristic curve at the X_{50} point. It is an index of the discrimination power of the item.

Table 3
Distribution of Fathers' Occupations

Occupation	Boys	Girls
PROFESSIONAL, TECHNICAL, AND KINDRED WORKERS		
00. Accountant	2	2
01. Architect	1	1
02. Dentist	--	--
03. Engineer	5	8
04. Lawyer, Judge	4	3
05. Clergyman	--	--
06. Doctor	7	4
07. Nurse	--	--
08. Teacher, Professor	18	21
09. Other Professional	16	22
FARMER		
11. Farmer	--	--
MANAGERS, OFFICIALS, PROPRIETORS, EXCEPT FARM		
21. Owner of Business	2	--
22. Manager, Official	12	11
CLERICAL AND KINDRED WORKERS		
31. Bookkeeper	--	--
32. Receptionist	--	--
39. Other Clerical and Kindred Workers	3	5
SALES WORKERS		
49. Salesman	20	15
CRAFTSMEN, FOREMEN, AND KINDRED WORKERS (SKILLED WORKERS)		
51. Craftsman, Skilled Worker	31	17
52. Foreman	2	4
53. Armed Services - Officer	1	1
54. Armed Services - Enlisted Man	1	--
OPERATIVES AND KINDRED WORKERS (SEMI-SKILLED WORKERS)		
61. Truck Driver	10	5
62. Operative in Factory	9	8
69. Other Operative and Kindred Workers	18	23
PRIVATE HOUSEHOLD AND SERVICE WORKERS		
71. Fireman	1	3
72. Policeman	1	--
73. Other Protective Service Worker	--	1
74. Practical Nurse, Nurse's Aide	2	--
75. Private Household Workers	1	--
79. Other Service Workers	14	13
81. Non-Farm Laborer	--	--
82. Farm Laborer	--	--
91. Not presently in labor force	4	8
99. Not ascertained	13	22

III
Results and Discussion

The means, standard deviations, and Hoyt reliability estimates obtained for the data collected during summer and fall of 1970 using the revised items are presented, separately for boys and girls, for total concept and total task scores. Also included in this section are a presentation and discussion of the item indices obtained for the correct choice of each item using both concept and task criterion scores.

Reliability Estimates and Test Statistics

Table 4 contains the means, standard deviations, and Hoyt reliability estimates obtained for the data collected during summer and fall, 1970, using the revised items for total concept and total task scores. The data were analyzed separately for the 195 boys and the 196 girls. The key for the task scores appears on the table; the key for the concept scores is given by the numbers in parentheses in the list of concepts presented in Table 1. For example, concept number 1 is Disjoint Sets, number 2 is Empty Sets, number 3 is Equal Sets, etc. In general, the concept scores consist of 12 items each, and the task scores of 30 items each. Exceptions to this are noted in two of the footnotes.

The mean scores for boys are generally lower than are the mean scores for girls. No conclusions can be drawn from this, however, as the data for the girls were collected in early summer shortly after the school year of their fifth grade had ended and the data for the boys were collected in the fall shortly after the school year of their sixth grade had begun. Thus, it cannot be determined what, if any, of this difference is due to a sex difference and what is due to a time difference and possible forgetting factor. It should also be noted that the scores for Concepts 8, 15, and 22 are based on one more item for boys than they are for girls; Concept 15 has 11 and 10 items for boys and girls respectively, Concepts 8 and 22 have 11 and 12 items respectively making up the total score. The scores for Tasks 1, 2, and 9 are made up of 30 items for boys but only 29 for girls.

The standard deviations and Hoyt reliability estimates are generally higher for boys than they are for girls.

The reliability estimates are sufficiently high to warrant study of the dimensionality of these selected mathematics concepts and the tasks when using mathematics content. This is a major objective of the CAA Project and is the main purpose for developing these items to measure mathematics concept attainment.

As was mentioned earlier, the subject matter specialists categorized the identified mathematics concepts into three major areas: Sets, Division, and Expressing Relationships. This was done on a theoretical basis. The data could be, and were, analyzed by area for task scores. Instead of a single total task score consisting of the score for that task type item for each of the 30 concepts, three different task scores were obtained for each of the 12 tasks, consisting of the score for that task type item for each of the 10 concepts within a single area. The mean, standard deviation, and Hoyt reliability estimate for each of these 36 scores, 3 areas by 12 tasks, were obtained. Table 5 contains the reliability estimates obtained for task scores by area and for the total across all 30 of the concepts. Spearman-Brown estimates for tripled test lengths (some are given at the bottom of Table 5 for comparison purposes) indicate that the area distinctions are not important ones; the reliability estimates for the total task scores are about what would be expected from tripling the length of the test when the single area reliability estimates are of the magnitude that were obtained. Also, pre-

Table 4
Means, Standard Deviations, and Reliabilities for
Mathematics Concept and Task Scores: Boys and Girls

	Concepts[a,b]						Tasks[c]					
	Mean		Standard Dev.		Hoyt Rel.		Mean		Standard Dev.		Hoyt Rel.	
No.	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1	6.51	7.18	2.20	2.01	.48	.42	18.89	19.34*	5.41	4.13	.81	.71
2	7.09	8.06	2.66	2.32	.67	.61	18.14	19.45*	5.72	5.12	.82	.80
3	6.15	7.25	2.61	2.41	.64	.62	20.16	22.25	5.04	4.00	.80	.73
4	6.99	7.42	2.34	2.34	.55	.60	20.42	22.79	5.03	4.00	.79	.73
5	7.51	8.34	2.33	2.01	.61	.49	18.52	21.05	5.56	4.34	.82	.75
6	6.95+	7.43+	1.99	1.80	.49	.41	16.79	19.44	6.21	5.56	.84	.82
7	5.48	6.36	2.54	2.48	.62	.62	12.63	12.51	4.93	4.61	.73	.70
8	6.82	6.59+	2.49	2.21	.62	.56	16.92	20.40	6.30	5.86	.85	.85
9	5.89	6.10	2.62	2.49	.63	.59	16.94	18.54*	6.02	5.37	.83	.81
10	6.63+	7.43+	2.63	2.04	.71	.58	15.28	17.16	5.33	5.11	.78	.78
11	6.68	8.10	3.00	2.53	.74	.66	11.85**	13.65**	4.49	3.73	.77	.68
12	7.18	8.57	2.66	2.24	.67	.61	12.25	13.51	4.13	3.99	.62	.58
13	5.02	5.48	2.58	2.55	.62	.60						
14	7.69	8.87	2.61	2.47	.69	.73						
15	7.14+	7.28++	2.51	2.14	.69	.68						
16	7.33+	7.64+	2.49	2.27	.71	.66						
17	6.26+	7.19+	2.39	2.28	.62	.63						
18	6.79	7.12	2.94	2.90	.75	.76						
19	6.20	6.74	2.69	2.51	.67	.64						
20	6.50	7.65	2.55	2.45	.64	.64						
21	5.66+	5.87+	2.19	2.09	.53	.52						
22	7.49	7.92+	2.29	1.71	.58	.42						
23	6.43	7.11	2.31	2.18	.57	.55						
24	5.21+	6.24+	2.31	2.29	.58	.60						
25	6.65	7.97	2.67	2.39	.65	.62						
26	5.65	6.32	2.58	2.16	.65	.50						
27	6.35	7.41	2.44	2.33	.61	.61						
28	6.83	7.42	2.58	2.02	.65	.44						
29	7.16	7.84	2.38	2.16	.59	.55						
30	8.55	9.21	2.52	1.93	.71	.64						

Key for Tasks:
1 Given name of attribute, select example.
2 Given example of attribute, select name.
3 Given name of concept, select example.
4 Given name of concept, select nonexample.
5 Given example of concept, select name.
6 Given concept, select relevant attribute.
7 Given concept, select irrelevant attribute.
8 Given definition of concept, select name.
9 Given name of concept, select definition.
10 Given concept, select supraordinate concept.
11 Given concept, select subordinate concept.
12 Given two concepts, select relationship.

[a] The key for the concepts is given by the numbers in parentheses in the list of concepts (Table 1).

[b] Scores consist of 12 items each except those marked as follows: + has 11 and ++ has 10.

[c] Scores consist of 30 items each except those marked as follows: * has 29 and ** has 23.

Table 5
Reliability Estimates for Task Scores by Area and Total for Girls

Task	Area			Total[b]
	Set Theory[a]	Division[a]	Expressing Relationships[a]	
1	.36	.45+	.51	.71*
2	.53+	.61	.57	.80*
3	.46	.53	.49	.73
4	.41	.55	.49	.73
5	.49	.59	.49	.75
6	.60	.65	.58	.82
7	.42	.54	.33	.70
8	.56	.73	.65	.85
9	.62	.63	.50+	.81*
10	.56	.66	.40	.78
11	.20++	.45+++	.48++	.68**
12	.26	.41	.19	.58

[a] Scores consist of 10 items each except those marked as follows: + has 9, ++ has 8, and +++ has 7.

[b] Scores consist of 30 items each except those marked as follows: * has 29 and ** has 23.

For comparison, these are the Spearman-Brown estimates for tripled test length:

Original	Estimated
.40	.67
.50	.75
.60	.82
.65	.85
.70	.88

liminary factor results indicate that the area distinctions are not important ones. The factor analyses of these data will be reported in a later paper.

Item Indices

Table 6 contains the item indices obtained, separately for boys and girls, based on both concept and task criterion scores. The indices included are proportion correct (this frequently is called difficulty or P), item-criterion biserial correlation, X_{50}, and β. They are given for the correct choice only. The key for the concepts is given by the numbers in parentheses in the list of concepts given in Table 1 (it is the same as for Table 4) and the key for the tasks is given in Table 4. The item number has no special meaning; it is a coding number and was included in the table as an organizational aid. Decimals have been omitted from the proportion correct and the biserial correlation columns. Note that proportion correct is the same whether analyzed using the concept criterion score or the task criterion score; hence, there is only one column each for boys and girls. The other item indices differ according to criterion score used. When an item was missing from the data collected, the appropriate row was left blank except for the identifying numbers, e.g., Item 203 for Concept 17 - Task 11. Three items, Nos. 71, 86, and 261, were missing from the data collected for the girls but were available for the boys; in this case only the columns for the girls are blank. There are a few instances where there is a blank in an X_{50} column. If β is very low, the X_{50} becomes essentially meaningless; thus, X_{50} is not included if the β value is less than .10.

Table 6
Item Indices Based on Concept and Task Criterion Scores

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
1 Disjoint Sets	1	1	68	79	46	61	59	58	-1.02	-.77	-1.38	-1.39	.52	.78	.72	.72
	2	2	85	94	57	55	53	62	-1.82	-1.88	-2.93	-2.48	.70	.67	.62	.80
	3	3	42	45	44	08	27	13	.45	2.44	.47	.97	.49	.08	.29	.13
	4	4	54	54	55	44	60	60	-.20	-.25	-.15	-.15	.66	.49	.76	.75
	5	5	47	60	59	55	59	43	.12	.13	-.41	-.57	.73	.65	.73	.48
	6	6	62	73	67	61	65	56	-.45	-.50	-.94	-1.09	.91	.78	.85	.67
	7	7	33	27	30	14	12	05	1.44	3.13	5.25		.31	.14	.12	.05
	8	8	59	69	61	66	58	61	-.37	-.34	-.88	-.83	.78	.89	.71	.78
	9	9	62	74	53	36	63	54	-.58	-.85	-1.04	-1.22	.62	.39	.82	.64
	10	10	30	36	52	38	58	47	.99	1.36	.63	.77	.61	.41	.72	.54
	11	11	65	62	47	26	40	22	-.82	-1.50	-.77	-1.44	.54	.27	.44	.22
	12	12	43	46	25	25	41	30	.76	.75	.25	.34	.25	.26	.45	.32
2 Empty Sets	1	13	45	55	56	52	58	62	.22	.24	-.22	-.21	.67	.61	.71	.78
	2	14	74	71	57	64	59	66	-1.13	-1.00	-.96	-.96	.69	.83	.73	.87
	3	15	74	82	79	72	69	62	-.81	-.89	-1.31	-1.44	1.27	1.04	.95	.80
	4	16	82	90	64	70	72	59	-1.43	-1.30	-1.77	-2.17	.84	.99	1.03	.72
	5	17	57	71	60	40	56	56	-.29	-.44	-.98	-.99	.74	.44	.68	.67
	6	18	49	65	64	53	63	70	.05	.06	-.60	-.54	.83	.63	.81	.99
	7	19	36	29	21	30	16	17	1.75	1.34	3.37	3.21	.21	.28	.17	.17
	8	20	70	84	70	64	76	66	-.76	-.82	-1.31	-1.51	.99	.84	1.18	.88
	9	21	59	73	66	63	75	70	-.37	-.38	-.84	-.89	.87	.82	1.12	.99
	10	22	64	72	65	49	61	56	-.55	-.74	-.97	-1.06	.86	.56	.78	.68
	11	23	53	57	60	44	41	22	-.14	-.19	-.40	-.76	.75	.49	.46	.23
	12	24	45	57	68	55	73	49	.18	.22	-.23	-.34	.93	.66	1.06	.56
3 Equal Sets	1	25	46	59	71	65	70	56	.15	.17	-.31	-.39	1.00	.85	.99	.67
	2	26	45	51	52	52	57	37	.26	.26	-.02	-.03	.61	.61	.69	.40
	3	27	49	52	44	30	53	40	.07	.11	-.10	-.13	.49	.31	.63	.43
	4	28	58	71	56	50	60	45	-.36	-.40	-.92	-1.24	.67	.58	.75	.50
	5	29	73	85	67	64	68	78	-.92	-.97	-1.51	-1.32	.91	.83	.92	1.24
	6	30	64	81	60	55	74	64	-.58	-.64	-1.20	-1.38	.75	.65	1.09	.83
	7	31	32	34	44	45	59	64	1.04	1.03	.72	.66	.49	.50	.73	.83
	8	32	71	85	83	68	77	80	-.89	-.82	-1.35	-1.31	.81	.93	1.21	1.33
	9	33	47	47	51	47	32	36	.14	.15	.20	.18	.59	.53	.34	.38
	10	34	62	80	74	69	73	68	-.40	-.43	-1.16	-1.24	1.09	.94	1.06	.94
	11	35	39	41	64	55	62	54	.43	.51	.35	.41	.84	.65	.79	.64
	12	36	30	39	47	40	40	35	1.12	1.34	.68	.78	.54	.43	.44	.37

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a] Boys[b]	Girls[c]	Biserial Correlation[a] Boys C	Boys T	Girls C	Girls T	X_{50} Boys C	Boys T	Girls C	Girls T	Beta Boys C	Boys T	Girls C	Girls T
4 Equivalent Sets	1	37	71	85	79	65	77	62	-.69	-.84	-1.33	-1.64	1.28	.86	1.21	.80
	2	38	43	51	63	57	73	72	.30	.33	-.03	-.04	.80	.69	1.08	1.03
	3	39	72	80	64	57	71	55	-.90	-1.00	-1.16	-1.49	.83	.70	1.02	.66
	4	40	64	67	56	52	63	54	-.63	-.67	-.71	-.83	.67	.61	.82	.65
	5	41	54	55	60	33	36	34	-.18	-.34	-.35	-.38	.76	.34	.39	.36
	6	42	67	74	65	56	58	56	-.69	-.80	-1.11	-1.15	.84	.67	.71	.67
	7	43	46	35	40	50	38	34	.27	.22	.99	1.12	.44	.58	.41	.36
	8	44	78	82	59	42	71	49	-1.34	-1.87	-1.26	-1.85	.73	.47	1.02	.56
	9	45	70	69	44	46	64	53	-1.17	-1.13	-.77	-.92	.49	.51	.83	.63
	10	46	32	35	24	17	48	26	1.99	2.81	.82	1.53	.24	.17	.55	.27
	11	47	62	70	58	58	61	60	-.53	-.53	-.87	-.89	.72	.71	.78	.75
	12	48	41	39	31	11	38	27	.72	2.01	.71	1.01	.33	.11	.42	.28
5 Line	1	49	68	61	40	20	42	09	-1.18	-2.37	-.64	-2.93	.44	.20	.47	.09
	2	50	72	75	63	49	53	22	-.94	-1.22	-1.28	-3.03	.81	.56	.62	.23
	3	51	91	92	51	52	71	78	-2.60	-2.54	-2.02	-1.84	.59	.61	1.00	1.24
	4	52	91	94	58	43	53	29	-2.34	-3.19	-2.93	-5.34	.71	.47	.62	.30
	5	53	68	69	55	44	70	57	-.84	-1.05	-.71	-.87	.66	.49	.97	.69
	6	54	67	83	69	57	74	73	-.65	-.78	-1.27	-1.28	.95	.69	1.10	1.08
	7	55	43	46	60	51	54	49	.29	.34	.17	.18	.75	.59	.63	.56
	8	56	53	69	60	34	51	35	-.14	-.25	-.97	-1.41	.74	.36	.59	.37
	9	57	65	73	68	57	45	48	-.55	-.66	-1.40	-1.30	.92	.70	.50	.55
	10	58	56	60	64	67	60	62	-.25	-.24	-.43	-.42	.82	.90	.74	.79
	11	59	56	75	60	56	45	31	-.27	-.29	-1.51	-2.21	.74	.67	.50	.32
	12	60	20	37	45	42	42	32	1.87	1.99	.81	1.07	.50	.47	.46	.34
6 Parallel Lines	1	61	36	40	51	45	53	40	.68	.77	.46	.62	.59	.51	.62	.43
	2	62	38	39	36	32	49	27	.86	.96	.58	1.06	.38	.34	.56	.28
	3	63	92	82	73	62	72	58	-1.94	-2.29	-1.25	-1.57	1.08	.80	1.03	.70
	4	64	82	92	75	80	70	64	-1.23	-1.15	-2.05	-1.02	1.13	1.32	.98	.78
	5	65	77	85	71	65	59	47	-1.06	-1.15	-1.76	-2.20	1.00	.86	.74	.54
	6	66	69	82	59	52	59	48	-.82	-.94	-1.53	-1.87	.74	.61	.73	.55
	7	67	46	53	36	42	60	42	.30	.26	-.13	-.18	.39	.47	.75	.46
	8	68	67	79	50	48	54	48	-.88	-.92	-1.50	-1.67	.58	.55	.64	.55
	9	69	81	85	82	69	79	73	-1.08	-1.27	-1.30	-1.41	1.41	.96	1.29	1.06
	10	70	56	59	57	51	46	42	-.29	-.32	-.50	-.56	.69	.59	.52	.46
	11	71														
	12	72	50	47	41	16	13	-10	.02	.04	.58	-.76	.44	.16	.13	-.10

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
7 Plane	1	73	52	57	59	42	56	37	-.08	-.11	-.32	-.49	.73	.46	.67	.39
	2	74	50	72	68	55	61	58	.01	.01	-.96	-1.00	.92	.66	.77	.71
	3	75	62	69	60	37	67	54	-.49	-.80	-.74	-.92	.74	.40	.90	.63
	4	76	55	69	60	45	80	52	-.20	-.27	-.61	-.95	.75	.50	1.34	.60
	5	77	63	74	63	59	67	57	-.51	-.54	-.95	-1.13	.81	.74	.91	.69
	6	78	47	55	53	48	45	26	.13	.15	-.26	-.44	.63	.55	.50	.27
	7	79	19	21	46	36	60	49	1.89	2.44	1.36	1.64	.52	.39	.74	.57
	8	80	53	52	64	59	56	41	-.13	-.14	-.09	-.13	.84	.73	.68	.45
	9	81	46	45	57	42	62	52	.17	.23	.18	.22	.69	.47	.80	.61
	10	82	46	58	44	51	53	35	.22	.19	-.39	-.60	.49	.60	.62	.37
	11	83	19	19	37	28	30	08	2.39	3.19	2.85		.40	.29	.32	.08
	12	84	36	44	59	54	49	34	.59	.64	.29	.42	.74	.65	.57	.36
8 Point	1	85	68	43	42	36	27	20	-1.13	-1.31	.62	.83	.46	.39	.28	.21
	2	86	79		56	65			-1.46	-1.27			.68	.85		
	3	84	69	83	72	55	70	52	-.70	-.92	-1.37	-1.85	1.04	.65	.99	.61
	4	88	43	47	47	32	43	26	.40	.60	.18	.30	.54	.33	.48	.26
	5	89	85	94	61	55	58	67	-1.71	-1.90	-2.73	-2.37	.77	.66	.72	.90
	6	90	39	39	31	12	35	30	.85	2.25	.82	.94	.33	.12	.37	.32
	7	91	55	61	45	41	62	49	-.30	-.33	-.46	-.58	.51	.45	.78	.57
	8	92	33	56	58	55	76	57	.76	.81	-.19	-.25	.72	.66	1.18	.70
	9	93	61	66	60	51	61	62	-.44	-.52	-.66	-.66	.75	.60	.78	.79
	10	94	50	52	74	58	62	60	.01	.01	-.08	-.08	1.12	.71	.80	.76
	11	95	48	62	79	72	69	59	.06	.06	-.43	-.51	1.29	1.04	.94	.73
	12	96	51	56	59	51	68	45	-.05	-.06	-.23	-.35	.73	.59	.94	.50
9 Subset	1	97	44	46	75	39	67	46	.22	.42	.13	.19	1.12	.42	.89	.52
	2	98	53	63	61	56	63	69	-.12	-.13	-.51	-.47	.76	.68	.82	.96
	3	99	43	37	70	35	59	36	.27	.53	.55	.91	.99	.38	.74	.38
	4	100	73	74	48	32	28	20	-1.30	-1.97	-2.36	-3.30	.55	.33	.29	.20
	5	101	46	38	59	36	52	15	.18	.30	.58	1.20	.74	.39	.61	.15
	6	102	46	53	42	45	59	60	.26	.24	-.13	-.13	.46	.51	.73	.75
	7	103	41	37	37	39	43	42	.66	.62	.78	.80	.39	.42	.48	.47
	8	104	48	56	56	42	56	36	.08	.11	-.28	-.43	.68	.47	.67	.39
	9	105	62	46	49	48	59	43	-.60	-.61	.15	.21	.56	.55	.74	.48
	10	106	46	49	64	36	43	41	.17	.31	.03	.03	.83	.38	.48	.45
	11	107	42	49	55	51	60	58	.39	.42	.02	.02	.66	.59	.75	.72
	12	108	48	59	55	41	50	44	.08	.11	-.46	-.53	.66	.45	.58	.49

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
10 Subtraction	1	109	58	46	52	45	53	33	-.39	-.45	.19	.31	.60	.50	.62	.35
	2	110	72	79	70	58	56	59	-.82	-.99	-1.42	-1.33	.98	.71	.67	.74
	3	111	71	91	78	63	71	79	-.72	-.89	-1.87	-1.68	1.24	.81	1.01	1.30
	4	112	74	90	58	64	75	59	-1.12	-1.02	-1.69	-2.17	.72	.84	1.14	.72
	5	113	73	94	80	77	81	70	-.76	-.79	-1.91	-2.22	1.34	1.20	1.38	.97
	6	114	56	69	70	68	72	63	-.21	-.22	-.71	-.80	.97	.92	1.03	.82
	7	115	36	34	51	68	60	66	.68	.51	.70	.64	.60	.92	.75	.87
	8	116	44	49	68	58	61	58	.24	.28	.02	.02	.94	.71	.76	.71
	9	117	58	67	79	65	72	62	-.26	-.31	-.60	-.70	1.27	.85	1.05	.79
	10	118	73	82	79	80	58	67	-.79	-.78	-1.57	-1.38	1.28	1.34	.72	.90
	11	119														
	12	120	47	42	39	31	46	39	.18	.23	.42	.50	.42	.33	.51	.42
11 Denominator	1	121	55	57	54	43	44	35	-.22	-.28	-.38	-.48	.65	.48	.49	.37
	2	122	55	67	44	38	47	43	-.31	-.36	-.96	-1.05	.49	.41	.53	.47
	3	123	62	77	87	63	71	60	-.34	-.46	-1.05	-1.23	1.76	.82	1.00	.75
	4	124	72	79	71	50	71	46	-.82	-1.15	-1.11	-1.72	1.00	.58	1.01	.52
	5	125	60	84	79	68	63	47	-.32	-.37	-1.56	-2.07	1.27	.93	.81	.54
	6	126	60	63	50	31	66	49	-.51	-.81	-.51	-.70	.58	.33	.89	.56
	7	127	48	60	66	50	66	45	.09	.12	-.39	-.58	.88	.58	.87	.50
	8	128	52	71	72	63	74	67	-.06	-.07	-.75	-.82	1.04	.82	1.09	.91
	9	129	62	79	72	46	78	57	-.41	-.63	-1.01	-1.40	1.04	.52	1.25	.69
	10	130	45	57	60	64	50	62	.23	.21	-.36	-.29	.74	.84	.58	.79
	11	131	54	70	61	53	78	65	-.16	-.18	-.67	-.80	.77	.62	1.24	.85
	12	132	45	46	54	52	38	42	.22	.24	.24	.22	.65	.61	.41	.46
12 Division	1	133	64	83	80	70	69	49	-.45	-.52	-1.39	-1.97	1.32	.98	.96	.56
	2	134	63	74	66	49	62	51	-.51	-.69	-1.04	-1.26	.88	.56	.78	.59
	3	135	83	94	61	58	77	65	-1.54	-1.61	-2.02	-2.36	.77	.72	1.19	.86
	4	136	80	89	72	69	58	75	-1.18	-1.22	-2.10	-1.62	1.03	.95	.71	1.14
	5	137	65	88	82	81	76	74	-.47	-.48	-1.54	-1.58	1.43	1.40	1.16	1.09
	6	138	46	62	53	62	66	68	.18	.16	-.47	-.46	.63	.78	.87	.92
	7	139	46	62	53	49	69	48	.18	.20	-.43	-.62	.62	.57	.95	.55
	8	140	58	64	59	52	70	63	-.34	-.39	-.52	-.58	.72	.60	.98	.81
	9	141	44	56	53	41	49	42	.30	.39	-.29	-.33	.63	.45	.55	.47
	10	142	70	74	52	44	71	63	-.99	-1.17	-.93	-1.04	.61	.49	1.01	.81
	11	143	53	60	64	51	33	32	-.11	-.14	-.78	-.80	.83	.59	.35	.34
	12	144	47	51	36	27	46	31	.23	.31	-.03	-.04	.39	.28	.51	.32

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
13 Factor	1	145	43	44	59	42	46	40	.30	.41	.31	.35	.73	.47	.51	.44
	2	146	36	43	58	55	47	34	.63	.66	.38	.52	.71	.66	.54	.37
	3	147	38	42	72	53	62	49	.43	.58	.33	.42	1.03	.62	.79	.56
	4	148	57	58	49	35	45	27	-.36	-.49	-.46	-.78	.56	.38	.50	.28
	5	149	49	49	60	58	59	46	.03	.03	.02	.03	.75	.72	.74	.51
	6	150	46	49	51	60	53	43	.19	.16	.02	.03	.59	.74	.62	.48
	7	151	38	37	39	03	44	35	.75		.73	.92	.43	.03	.50	.38
	8	152	48	65	76	67	70	74	.08	.09	-.56	-.53	1.16	.90	.99	1.11
	9	153	35	37	56	26	61	43	.67	1.46	.56	.79	.68	.27	.77	.47
	10	154	42	43	35	45	41	45	.61	.47	.44	.40	.37	.51	.45	.50
	11	155	33	42	67	46	73	31	.66	.97	.27	.62	.91	.52	1.06	.33
	12	156	37	37	47	37	54	46	.71	.90	.60	.70	.53	.40	.65	.52
14 Fraction	1	157	68	76	61	46	64	41	-.78	-1.02	-1.10	-1.73	.77	.52	.84	.45
	2	158	62	73	67	54	84	68	-.44	-.54	-.72	-.90	.89	.65	1.57	.93
	3	159	88	96	76	66	73	70	-1.53	-1.75	-2.38	-2.49	1.16	.89	1.07	.98
	4	160	87	93	64	46	95	76	-1.74	-2.42	-1.58	-1.97	.83	.52	3.11	1.19
	5	161	85	94	45	51	69	56	-2.32	-2.06	-2.24	-2.76	.50	.59	.95	.67
	6	162	50	70	72	69	67	52	.01	.01	-.80	-1.02	1.04	.95	.91	.62
	7	163	40	50	52	37	60	33	.49	.69	.00	.00	.61	.40	.75	.35
	8	164	65	81	74	62	83	61	-.51	-.61	-1.04	-1.42	1.10	.78	1.48	.76
	9	165	69	76	65	62	74	77	-.77	-.81	-.96	-.92	.85	.79	1.09	1.20
	10	166	57	55	62	52	64	52	-.28	-.33	-.20	-.25	.79	.61	.84	.61
	11	167	56	73	74	67	74	65	-.20	-.22	-.82	-.95	1.09	.89	1.11	.85
	12	168	43	49	53	51	69	61	.33	.34	.02	.02	.62	.59	.95	.76
15 Mixed Fraction	1	169	83		64	67			-1.50	-1.44			.83	.89		
	2	170	66	77	69	73	78	72	-.58	-.55	-.95	-1.03	.96	1.06	1.26	1.03
	3	171	77	92	76	67	78	47	-.97	-1.10	-1.78	-2.95	1.15	.90	1.25	.54
	4	172	84	91	63	61	72	67	-1.56	-1.60	-1.84	-2.00	.80	.77	1.05	.89
	5	173	67	84	68	62	83	66	-.65	-.72	-1.18	-1.50	.93	.79	1.50	.87
	6	174	64	68	71	61	78	63	-.51	-.59	-.61	-.76	1.01	.77	1.24	.81
	7	175	43	48	47	54	52	42	.37	.32	.10	.12	.53	.65	.62	.46
	8	176	65	74	67	64	85	76	-.56	-.58	-.78	-.87	.90	.84	1.59	1.16
	9	177	59	74	75	65	86	68	-.32	-.37	-.76	-.97	1.14	.85	1.69	.92
	10	178	62	76	68	51	75	70	-.45	-.60	-.92	-.98	.93	.59	1.15	.99
	11	179														
	12	180	45	44	50	44	45	37	.27	.30	.34	.41	.58	.50	.51	.40

GPO 826-816-4

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
16 Multiplication	1	181	46	49	82	68	84	57	.12	.14	.03	.05	1.45	.92	1.53	.69
	2	182	61	65	77	51	79	46	-.35	-.53	-.50	-.86	1.22	.59	1.29	.51
	3	183	87	93	57	52	45	53	-1.95	-2.15	-3.29	-2.77	.69	.60	.50	.62
	4	184	82	89	61	61	45	52	-1.50	-1.50	-2.70	-2.35	.77	.77	.50	.60
	5	185	87	90	74	55	75	79	-1.50	-2.00	-1.73	-1.65	1.11	.67	1.13	1.28
	6	186	66	73	67	62	71	62	-.60	-.65	-.88	-1.01	.91	.78	1.01	.79
	7	187	52	44	58	42	50	41	-.08	-.11	.28	.34	.72	.46	.58	.45
	8	188	68	67	64	48	66	25	-.72	-.95	-.66	-1.76	.84	.55	.87	.26
	9	189	67	66	77	61	72	58	-.56	-.71	-.56	-.70	1.22	.76	1.05	.72
	10	190	75	79	62	65	60	62	-1.03	-1.04	-1.32	-1.28	.79	.84	.75	.79
	11	191														
	12	192	45	49	57	47	62	36	.24	.29	.04	.07	.70	.53	.78	.38
17 Numerator	1	193	68	66	59	59	44	36	-.77	-.77	-.96	-1.17	.74	.74	.49	.39
	2	194	58	65	61	57	48	48	-.33	-.35	-.82	-.82	.77	.69	.54	.55
	3	195	72	79	62	44	70	50	-.95	-1.34	-1.16	-1.63	.80	.49	.98	.57
	4	196	66	73	55	48	76	61	-.73	-.83	-.83	-1.02	.66	.55	1.16	.78
	5	197	74	83	51	42	74	53	-1.27	-1.58	-1.28	-1.77	.60	.46	1.09	.63
	6	198	61	79	65	56	64	41	-.43	-.50	-1.27	-1.96	.85	.67	.83	.45
	7	199	42	36	59	61	59	74	.34	.33	.62	.50	.73	.78	.74	1.09
	8	200	58	73	78	70	78	72	-.27	-.31	-.81	-.87	1.26	.98	1.24	1.04
	9	201	59	76	65	62	80	44	-.37	-.46	-.88	-1.59	.85	.61	1.33	.50
	10	202	35	47	54	48	53	56	.72	.80	.14	.14	.64	.55	.63	.67
	11	203														
	12	204	32	41	31	23	44	24	1.46	1.97	.52	.97	.33	.24	.49	.25
18 Product	1	205	81	90	68	67	52	54	-1.29	-1.30	-2.52	-2.39	.94	.91	.60	.65
	2	206	75	79	67	57	60	68	-1.00	-1.18	-1.31	-1.17	.91	.69	.76	.93
	3	207	45	51	68	52	77	49	.20	.26	-.02	-.03	.92	.61	1.21	.56
	4	208	58	71	74	58	67	62	-.27	-.34	-.82	-.89	1.09	.72	.91	.79
	5	209	62	49	66	44	70	46	-.47	-.69	.02	.03	.87	.50	.98	.52
	6	210	73	68	63	70	73	65	-.99	-.89	-.65	-.74	.81	.99	1.08	.85
	7	211	34	28	52	53	56	56	.77	.75	1.07	1.07	.61	.63	.67	.67
	8	212	51	52	71	61	76	61	-.04	-.05	-.07	-.08	1.02	.76	1.17	.77
	9	213	47	54	74	62	72	39	.11	.13	-.13	-.23	1.09	.81	1.03	.43
	10	214	59	74	70	71	84	76	-.32	-.32	-.79	-.87	.98	1.00	1.54	1.17
	11	215	54	48	50	49	56	62	-.19	-.20	.07	.06	.58	.55	.67	.78
	12	216	40	47	70	56	69	58	.36	.45	.09	.11	.97	.68	.95	.71

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
19 Quotient	1	217	85	91	52	52	48	52	-2.02	-1.99	-2.82	-2.61	.60	.61	.55	.61
	2	218	62	74	56	54	63	55	-.53	-.54	-1.04	-1.20	.67	.64	.82	.65
	3	219	54	55	65	48	60	36	-.15	-.20	-.19	-.32	.86	.55	.75	.39
	4	220	37	35	60	39	58	41	.55	.87	.68	.97	.76	.42	.72	.44
	5	221	43	40	58	47	72	54	.30	.37	.36	.48	.72	.53	1.03	.64
	6	222	60	73	77	61	76	65	-.33	-.42	-.83	-.96	1.22	.77	1.17	.86
	7	223	53	56	51	45	47	30	-.14	-.16	-.30	-.47	.60	.51	.53	.32
	8	224	42	56	64	49	76	66	.33	.43	-.20	-.23	.84	.57	1.15	.88
	9	225	54	54	76	61	76	60	-.14	-.18	-.14	-.17	1.16	.76	1.15	.75
	10	226	60	65	69	63	52	52	-.37	-.40	-.76	-.76	.95	.81	.61	.61
	11	227	39	38	38	52	34	44	.74	.54	.88	.68	.41	.60	.36	.49
	12	228	32	36	45	34	37	35	1.06	1.41	.96	1.00	.50	.36	.39	.38
20 Remainder	1	229	39	46	70	68	64	64	.38	.39	.14	.14	.97	.92	.84	.84
	2	230	54	55	68	63	66	58	-.14	-.15	-.18	-.20	.93	.80	.87	.70
	3	231	66	80	64	71	67	65	-.65	-.59	-1.27	-1.31	.83	1.00	.89	.85
	4	232	84	87	52	48	45	60	-1.87	-2.05	-2.45	-1.87	.61	.54	.51	.74
	5	233	71	80	67	54	64	46	-.84	-1.04	-1.31	-1.83	.90	.64	.84	.52
	6	234	39	55	56	36	64	45	.50	.79	-.18	-.26	.67	.38	.83	.50
	7	235	57	62	66	48	54	47	-.28	-.39	-.58	-.67	.88	.55	.65	.53
	8	236	61	78	61	62	86	77	-.46	-.45	-.88	-.98	.77	.79	1.68	1.22
	9	237	67	81	65	71	67	58	-.66	-.61	-1.32	-1.51	.86	1.01	.90	.72
	10	238	30	44	30	10	47	19	1.80	5.22	.33	.81	.31	.10	.53	.19
	11	239	51	61	47	53	59	44	-.07	-.06	-.48	-.64	.54	.62	.73	.50
	12	240	31	36	47	23	40	28	1.06	2.22	.88	1.24	.54	.23	.44	.30
21 Average	1	241	87	95	68	56	48	76	-1.64	-1.98	-3.40	-2.16	.92	.68	.55	1.16
	2	242	59	63	56	27	56	45	-.43	-.89	-.61	-.75	.68	.28	.67	.51
	3	243	40	37	73	63	66	34	.34	.40	.51	1.00	1.08	.82	.89	.36
	4	244	46	46	50	32	45	20	.22	.34	.23	.51	.58	.34	.50	.20
	5	245	36	40	48	27	75	39	.72	1.28	.35	.67	.55	.28	1.13	.42
	6	246	59	65	56	56	57	40	-.43	-.43	-.67	-.94	.67	.68	.69	.44
	7	247	37	31	33	33	44	53	.99	.96	1.14	.96	.34	.35	.50	.62
	8	248	52	63	65	56	71	57	-.09	-.10	-.48	-.59	.87	.67	1.00	.70
	9	249	52	55	64	50	73	53	-.09	-.12	-.16	-.22	.83	.57	1.08	.63
	10	250	27	20	36	15	25	10	1.69	3.95	3.28	8.48	.38	.16	.26	.10
	11	251														
	12	252	69	72	52	36	33	36	-.96	-1.39	-1.82	-1.65	.61	.39	.35	.39

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
22 Graph	1	253	80	89	47	41	51	56	-1.78	-2.04	-2.45	-2.21	.54	.45	.59	.68
	2	254	41	51	43	25	61	42	.53	.90	-.04	-.06	.48	.26	.77	.46
	3	255	91	94	83	58	35	14	-1.65	-2.36	-4.59	-11.04	1.46	.70	.37	.15
	4	256	70	81	64	47	55	32	-.83	-1.13	-1.60	-2.79	.84	.54	.66	.33
	5	257	83	93	60	60	68	64	-1.60	-1.59	-2.17	-2.28	.75	.76	.92	.84
	6	258	71	81	64	56	77	65	-.85	-.98	-1.12	-1.32	.83	.67	1.21	.86
	7	259	47	42	46	36	49	16	.15	.19	.42	1.26	.52	.39	.56	.17
	8	260	43	64	61	55	62	57	.31	.34	-.57	-.62	.78	.67	.78	.69
	9	261	52		47	46			-.12	-.12			.53	.52		
	10	262	32	36	54	32	43	23	.85	1.42	.85	1.57	.65	.34	.48	.24
	11	263	69	86	56	64	53	59	-.89	-.78	-2.03	-1.81	.68	.84	.62	.73
	12	264	69	76	70	62	54	38	-.71	-.81	-1.27	-1.80	.99	.80	.65	.42
23 Measurement	1	265	27	19	29	14	48	20	2.16	4.54	1.78	4.25	.30	.14	.55	.21
	2	266	57	59	54	49	52	43	-.32	-.36	-.42	-.51	.65	.56	.62	.47
	3	267	86	93	57	49	63	56	-1.87	-2.16	-2.33	-2.63	.69	.57	.81	.67
	4	268	79	90	53	55	66	62	-1.54	-1.49	-1.92	-2.05	.63	.66	.88	.79
	5	269	32	40	28	18	20	13	1.68	2.62	1.32	2.00	.29	.18	.20	.13
	6	270	56	64	65	64	64	52	-.25	-.25	-.55	-.67	.85	.83	.83	.62
	7	271	34	33	61	56	58	33	.58	.30	.75	1.33	.77	.70	.71	.35
	8	272	62	71	51	21	57	53	-.57	.83	-.97	-1.04	.60	.21	.69	.63
	9	273	43	55	58	35	66	59	.30	-.53	-.19	-.22	.71	.38	.89	.74
	10	274	50	55	69	58	60	55	.01	.01	-.21	-.23	.94	.71	.75	.66
	11	275	80	83	77	59	56	59	-1.10	-1.42	-1.69	-1.59	1.20	.74	.67	.74
	12	276	38	50	57	44	66	54	.54	.70	.00	.00	.69	.48	.87	.64
24 Open Sentence	1	277	70	87	63	69	66	55	-.84	-.77	-1.70	-2.04	.82	.96	.87	.65
	2	278	66	63	66	60	71	72	-.63	-.69	-.48	-.47	.88	.75	1.01	1.05
	3	279	49	55	50	33	50	19	.06	.10	-.23	-.62	.58	.35	.58	.19
	4	280	45	62	50	42	57	41	.25	.29	-.52	-.72	.58	.47	.70	.46
	5	281	42	54	72	65	74	48	.28	.31	-.12	-.19	1.03	.85	1.09	.55
	6	282	61	69	44	30	47	36	-.60	-.90	-1.05	-1.38	.49	.31	.53	.38
	7	283	34	33	47	41	42	20	.85	.98	1.06	2.23	.54	.45	.47	.21
	8	284	53	77	74	73	75	70	-.10	-.10	-.99	-1.06	1.09	1.07	1.12	.98
	9	285	37	43	59	39	68	33	.54	.81	.24	.50	.73	.43	.93	.36
	10	286	42	54	51	46	59	59	.40	.44	-.15	-.15	.59	.52	.74	.73
	11	287														
	12	288	21	29	35	17	34	10	2.30	4.69	1.68	5.85	.37	.17	.36	.10

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a] Boys[b]	Girls[c]	Biserial Correlation[a] Boys C	Boys T	Girls C	Girls T	X_{50} Boys C	Boys T	Girls C	Girls T	Beta Boys C	Boys T	Girls C	Girls T
25 Place Holder	1	289	62	57	61	66	62	69	-.51	-.46	-.27	-.24	.77	.88	.78	.96
	2	290	49	61	66	53	50	41	.05	-.08	-.57	-.70	.87	.62	.58	.45
	3	291	57	74	65	53	67	54	-.27	-.33	-.99	-1.21	.87	.62	.90	.65
	4	292	55	65	49	39	76	58	-.28	-.35	-.50	-.65	.55	.42	1.19	.71
	5	293	62	84	54	38	55	52	-.54	-.77	-1.83	-1.92	.65	.41	.66	.61
	6	294	65	69	70	75	65	66	-.54	-.50	-.78	-.77	.98	1.12	.85	.88
	7	295	55	64	47	41	37	39	-.26	-.30	-.99	-.93	.53	.45	.40	.43
	8	296	62	86	66	70	60	66	-.47	-.44	-1.81	-1.64	.87	.98	.76	.89
	9	297	61	82	64	51	66	57	-.42	-.52	-1.37	-1.57	.82	.59	.87	.70
	10	298	63	66	40	19	42	23	-.80	-1.71	-1.02	-1.80	.43	.19	.46	.24
	11	299	33	37	65	64	79	65	.66	.67	.43	.52	.85	.85	1.28	.85
	12	300	43	51	45	41	48	44	.42	.46	-.05	-.06	.50	.45	.55	.49
26 Place Value	1	301	69	78	68	56	65	62	-.72	-.87	-1.19	-1.26	.93	.68	.86	.78
	2	302	39	46	64	47	40	45	.41	.56	.26	.23	.84	.54	.43	.50
	3	303	70	72	67	54	76	70	-.79	-.98	-.79	-.86	.91	.65	1.16	.97
	4	304	70	87	62	63	42	53	-.83	-.82	-2.65	-2.10	.79	.81	.46	.63
	5	305	49	48	53	47	54	43	.04	.04	.09	.12	.63	.53	.64	.48
	6	306	38	40	60	58	54	49	.51	.52	.48	.53	.76	.72	.64	.56
	7	307	38	37	47	49	47	45	.62	.60	.70	.72	.54	.57	.53	.51
	8	308	38	47	60	40	52	51	.51	.77	.15	.15	.76	.43	.61	.59
	9	309	47	59	69	61	60	50	.12	.14	-.37	-.44	.95	.76	.75	.57
	10	310	22	17	35	15	30	23	2.25	5.30	3.24	4.26	.37	.15	.31	.23
	11	311	53	71	76	71	59	51	-.11	-.12	-.94	-1.09	1.17	1.01	.72	.59
	12	312	32	30	28	24	40	20	1.68	1.95	1.30	2.57	.29	.25	.44	.21
27 Solution Set	1	313	61	64	56	39	55	43	-.50	-.71	-.67	-.86	.68	.43	.66	.47
	2	314	66	81	66	64	72	72	-.71	-.74	-1.22	-1.22	.89	.82	1.05	1.05
	3	315	92	97	54	50	41	71	-2.63	-2.87	-4.79	-2.74	.65	.57	.45	1.02
	4	316	68	78	70	67	66	65	-.65	-.68	-1.18	-1.19	.99	.91	.87	.86
	5	317	43	56	59	55	63	48	.29	.32	-.22	-.30	.74	.65	.81	.54
	6	318	44	52	72	65	72	71	.22	.25	-.05	-.05	1.05	.85	1.04	1.02
	7	319	30	42	16	16	19	26	3.29	3.28	1.06	.79	.16	.16	.20	.27
	8	320	57	79	65	55	69	66	-.27	-.32	-1.14	-1.20	.86	.66	.96	.88
	9	321	49	62	65	62	73	57	.03	.03	-.43	-.55	.86	.79	1.06	.69
	10	322	53	54	57	48	64	48	-.15	-.17	-.16	-.21	.70	.55	.83	.55
	11	323	45	55	48	36	51	41	.28	.38	-.23	-.28	.55	.39	.59	.45
	12	324	25	21	44	42	52	28	1.53	1.61	1.53	2.88	.49	.46	.61	.29

Table 6. (Continued)

Concept	Task	Item	Proportion Correct[a]		Biserial Correlation[a]				X_{50}				Beta			
					Boys		Girls		Boys		Girls		Boys		Girls	
			Boys[b]	Girls[c]	C	T	C	T	C	T	C	T	C	T	C	T
28 Standard Unit	1	325	73	91	58	69	58	30	-1.08	-.90	-2.30	-4.43	.70	.96	.71	.31
	2	326	75	78	63	64	66	54	-1.06	-1.04	-1.17	-1.44	.82	.84	.88	.64
	3	327	65	76	66	71	59	64	-.57	-.53	-1.17	-1.08	.88	1.00	.73	.84
	4	328	76	82	65	59	50	62	-1.10	-1.21	-1.81	-1.45	.86	.74	.57	.79
	5	329	47	57	67	63	58	46	.12	.13	-.29	-.36	.91	.81	.72	.52
	6	330	42	31	40	15	22	-02	.53	1.38	2.21		.44	.16	.23	-.02
	7	331	55	51	59	42	42	35	-.21	-.29	-.03	-.04	.73	.46	.46	.38
	8	332	59	66	70	56	70	69	-.32	-.40	-.60	-.61	.99	.68	.98	.95
	9	333	50	57	60	53	52	58	.01	.01	-.35	-.31	.76	.63	.61	.71
	10	334	50	52	57	35	53	47	.01	.02	-.10	-.11	.68	.38	.62	.53
	11	335	66	74	66	49	54	52	-.63	-.85	-1.19	-1.25	.88	.56	.65	.60
	12	336	26	29	28	20	23	05	2.29	3.13	2.43		.29	.21	.24	.05
29 Statement	1	337	87	85	62	60	68	76	-1.79	-1.84	-1.50	-1.35	.79	.76	.93	1.16
	2	338	83	91	69	63	75	69	-1.40	-1.52	-1.81	-1.98	.94	.81	1.14	.95
	3	339	50	66	62	57	61	57	-.01	-.01	-.67	-.71	.80	.70	.77	.70
	4	340	69	86	62	49	60	48	-.81	-1.03	-1.77	-2.22	.80	.56	.76	.55
	5	341	72	75	72	66	69	67	-.80	-.87	-.98	-1.00	1.04	.89	.94	.91
	6	342	47	69	72	70	74	74	.10	.10	-.67	-.67	1.03	.99	1.09	1.09
	7	343	43	29	59	57	41	58	.30	.30	1.33	.95	.73	.70	.45	.71
	8	344	43	40	32	21	35	20	.55	.83	.71	1.22	.33	.21	.37	.20
	9	345	57	56	51	35	49	24	-.37	-.53	-.29	-.58	.60	.38	.57	.25
	10	346	72	81	64	49	62	57	-.90	-1.19	-1.39	-1.51	.84	.56	.79	.69
	11	347	36	45	39	38	46	35	.88	.91	.25	.33	.43	.41	.52	.37
	12	348	56	62	36	32	54	51	-.41	-.46	-.55	-.59	.38	.34	.64	.59
30 Weight	1	349	85	94	83	66	82	61	-1.25	-1.58	-1.89	-2.52	1.50	.87	1.42	.76
	2	350	76	87	56	47	88	64	-1.26	-1.49	-1.27	-1.75	.68	.53	1.84	.83
	3	351	82	91	74	72	83	73	-1.24	-1.27	-1.60	-1.82	1.10	1.04	1.43	1.07
	4	352	82	91	82	75	75	61	-1.09	-1.20	-1.78	-2.17	1.44	1.12	1.13	.77
	5	353	79	93	84	66	96	77	-.96	-1.22	-1.57	-1.95	1.56	.89	3.32	1.22
	6	354	66	70	65	64	56	47	-.64	-.65	-.93	-1.11	.86	.84	.68	.53
	7	355	48	30	60	50	44	44	.10	.12	1.17	1.20	.75	.58	.50	.48
	8	356	79	84	69	60	93	85	-1.19	-1.36	-1.06	-1.15	.96	.76	2.50	1.62
	9	357	72	86	65	65	74	59	-.91	-.92	-1.48	-1.84	.85	.85	1.10	.74
	10	358	69	83	76	72	75	44	-.66	-.70	-1.25	-2.13	1.18	1.04	1.14	.49
	11	359	78	86	80	58	67	49	-.96	-1.34	-1.61	-2.17	1.33	.71	.89	.57
	12	360	38	27	23	24	50	39	1.26	1.25	1.22	1.56	.24	.24	.58	.42

[a] Decimals have been omitted. [b] $N = 195$ [c] $N = 196$

If desired, the items that make up a criterion score can be separated out. This is easy to do for a concept; the items composing the criterion score are simply the 12 given in order consisting of one of each task type. For example, the items composing the criterion score for Concept 3 are numbered 25 through 36. The items composing the criterion score for a task are those with the same task number for each of the concepts; for example, the items composing the criterion score for Task 1 are numbered 1, 13, 25, 37, etc., with the last one being number 349.

As was evident from the means of the total scores, and as can be seen from the two difficulty indices given for the items (proportion correct and X_{50}), the items, in general, were more difficult for the boys than for the girls. There is not a one-to-one correspondence for each item, however; there are some exceptions, since some items were more difficult for the girls and some were about the same. As was pointed out earlier, however, no conclusions can be drawn from this because the data for the girls were collected in early summer shortly after the end of their fifth grade school year and the data for the boys were collected in the fall shortly after their sixth grade year had begun. The difficulty indices obtained indicate that these items are of appropriate difficulty levels for these subjects.

It seems clear from looking at Table 6 that X_{50} gives more precise information about the difficulty level of an item when that same item is a part of each of two criterion scores. The proportion correct remains the same for both of the criterion scores. This index tells how many subjects responded to the correct answer for an item but it says nothing about their ability level as measured by a particular criterion score--total concept score or total task score in this case. The item difficulty index, X_{50}, gives (in standard deviation units) the criterion score at which a subject would have a 50-50 chance of getting the item correct. For example, an X_{50} value of 1.20 for an item indicates that subjects with a criterion score 1.20 standard deviation units <u>above</u> the mean have a 50% chance of answering that item correctly. Subjects with a criterion score higher than this would have a greater chance of answering that item correctly, and subjects with a criterion score lower than this would have a lesser chance. Likewise, an X_{50} value of -1.20 means that subjects with a criterion score 1.20 standard deviation units <u>below</u> the mean would have a 50% chance of getting that item correct; for a higher score the chance would be greater, and for a lower score the chance would be less. Knowing both X_{50} and β for an item allows one to readily determine the probability of answering an item correctly for any point on the criterion scale (Baker, 1964). It may be pointed out that when P = .50, X_{50} = .00; when P is greater than .50 then X_{50} will be negative and, for a certain P, the higher the β value the closer to zero will be the X_{50} value. This can be seen from inspecting Table 6. For example, for Item 1 the β is higher for the concept score than it is for the task score for both boys and girls; similarly for both boys and girls, the X_{50} value is closer to zero for the concept score than it is for the task score. For P less than .50, the X_{50} will be positive, and again, for a certain P, the higher the β value the closer to zero will be the X_{50} value. See item 7 for an illustration of this.

The two item discrimination indices, biserial correlation and β, are more closely related since β is computed as a function of the biserial correlation (Baker, 1969). They are not linearly related, however. From .00 to about .30 (absolute) they are very nearly the same; beyond this, β begins to increase quite rapidly in magnitude. It may be pointed out that β is always equal to or greater (absolute) than the biserial correlation. As a general rule, .30 is often used as a lower cutting point for a desirable biserial correlation or β. For a total score composed of relatively few items, as is the concept score, a much higher minimum would be desirable.

As can be seen from Table 6, most of the mathematics items have desirable biserial correlations and βs when the item is both a part of a concept criterion score and a task criterion score. The most obvious thing is that the βs are higher, with a few exceptions, when the item is a part of a concept criterion score than when it is a part of a task criterion score. This is to be expected since the concept score consists of considerably fewer items than does the task score--12 items for most concept scores and 30 items for most task scores. The item-criterion biserial correlation is a part-whole correlation, with the criterion the total score of which the item is a part, and the fewer the number of items the greater should be the correlation of that item with the total score of which it is a part. Since β is computed as a function of the biserial correlation, it is affected in the same manner. There does not seem to be a consistent pattern in the magnitude of the βs for the boys as compared with the girls. For some of the items, the βs are considerably higher for the boys and for some of them they are considerably higher

for the girls. For the tryouts of the items, data for both boys and girls were analyzed together. If the data for boys and girls were pooled and item analyzed, the β values would probably increase for most of the items.

As was discussed earlier, these item indices were obtained by performing conventional item analyses on two different types of scores—one for concept criterion scores and one for task criterion scores. This was necessitated by the lack of item analysis procedures appropriate for use with data collected using a completely crossed design to build the items. It is not known how the item indices would be affected if procedures were available to compute them simultaneously taking into account the effects of the crossed design. A guess would be that discrimination indices would be affected more than would difficulty indices, if there were an effect. It is plausible to expect that there may be some concept-task interactions which cannot be, at least readily, ascertained by doing a conventional item analysis on the two types of scores.

GPO 826-616-3

IV
Summary and Conclusions

The primary objective of the project entitled "A Structure of Concept Attainment Abilities" is to formulate one or more models or structures of concept attainment abilities, and to assess their consistency with actual data. One of the major steps for attaining this primary objective was taken to be the development of tests to measure achievement of selected language arts, mathematics, science, and social studies concepts appropriate at the fourth grade level. This paper describes the test development efforts and presents the item and total score statistics obtained using the revised items developed for measuring achievement of selected concepts in mathematics.

Subject matter specialists identified single- or compound-word classificatory concepts for three major areas, and randomly selected 10 from each area to be studied. These 30 selected concepts were then analyzed. Twelve items for each concept were developed; one for each of the first 12 tasks of "A Schema for Testing Level of Concept Mastery" (Frayer, Fredrick, & Klausmeier, 1969).

The items that were developed were administered during early summer of 1970 to 196 girls who had just completed the fifth grade and during the fall of 1970 to 195 boys who had just begun the sixth grade. These data were item analyzed, separately for boys and for girls, using the GITAP program (Baker, 1969).

The means, standard deviations, and Hoyt reliability estimates obtained are presented and discussed for total concept and total task scores. Four different item indices--proportion correct, item-criterion biserial correlation, X_{50}, and β--obtained for each item based on each of two criterion scores, appropriate total concept score and appropriate total task score, are presented and discussed.

Conclusions

The major conclusions drawn are:

1. The reliability estimates obtained for both total concept scores and total task scores are sufficiently high to warrant study of the dimensionality of these selected mathematics concepts and the dimensionality of the tasks when using mathematics content.

2. The three area distinctions seem not to be important ones.

3. The difficulty item indices obtained indicate that these items are of appropriate difficulty levels for these subjects.

4. Most of the items have desirable levels of discrimination indices when the item is both a part of a concept criterion score and a task criterion score.

Recommendation

The completely crossed design used to construct these achievement tests is a very interesting one. This type of design might well be used more often in the future. It would be highly desirable to have available item analysis procedures that are appropriate for analyzing such crossed designs. At the present such a methodology is not known.

References

Baker, F. B. An intersection of test score interpretation and item analysis. Journal of Educational Measurement, 1964, 1, 23-28.

Baker, F. B. FORTAP: a fortran test analysis package. Department of Educational Psychology, The University of Wisconsin, 1969.

Bourne, L. E., Jr. Human conceptual behavior. Boston: Allyn & Bacon, 1966.

Frayer, D. A., Fredrick, W. C., and Klausmeier, H. J. A schema for testing the level of concept mastery. Wisconsin Research and Development Center for Cognitive Learning, Working Paper No. 16, 1969.

Harris, M. L., Harris, C. W., Frayer, D. A., and Quilling, M. R. A structure of concept attainment abilities: the problem and strategies for attacking it. Wisconsin Research and Development Center for Cognitive Learning, Theoretical Paper No. 32, in press.

Klausmeier, H. J., Harris, C. W., Davis, J. K., Schwenn, E., and Frayer, D. Strategies and cognitive processes in concept learning. The University of Wisconsin, Cooperative Research Project No. 2850, 1968.

Romberg, T. A. & Steitz, J. Items to test level of attainment of mathematics concepts by intermediate-grade children. Wisconsin Research and Development Center for Cognitive Learning, Working Paper No. 56, in press.

Romberg, T. A., Steitz, J., & Frayer, D. Selection and analysis of mathematics concepts for inclusion in tests of concept attainment. Wisconsin Research and Development Center for Cognitive Learning, Working Paper No. 55, in press.

GPO 926-616-2

National Evaluation Committee

Helen Bain
Immediate Past President
National Education Association

Lyle E. Bourne, Jr.
Institute for the Study of Intellectual Behavior
Universiy of Colorado

Jeanne S. Chall
Graduate School of Education
Harvard University

Francis S. Chase
Department of Education
University of Chicago

George E. Dickson
College of Education
University of Toledo

Hugh J. Scott
Superintendent of Public Schools
District of Columbia

H. Craig Sipe
Department of Instruction
State University of New York

G. Wesley Sowards
Dean of Education
Florida International University

Benton J. Underwood
Department of Psychology
Northwestern University

Robert J. Wisner
Mathematics Department
New Mexico State University

Executive Committee

William R. Bush
Director of Program Planning and Management
and Deputy Director, R & D Center

Herbert J. Klausmeier, Committee Chairman
Director, R & D Center

Wayne Otto
Principal Investigator
R & D Center

Robert G. Petzold
Professor of Music
University of Wisconsin

Richard A. Rossmiller
Professor of Educational Administration
University of Wisconsin

James E. Walter
Coordinator of Program Planning
R & D Center

Russell S. Way, ex officio
Program Administrator, Title III ESEA
Wisconsin Department of Public Instruction

Faculty of Principal Investigators

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor
Educational Psychology

Marvin J. Fruth
Associate Professor
Educational Administration

John G. Harvey
Associate Professor
Mathematics

Frank H. Hooper
Associate Professor
Child Development

Herbert J. Klausmeier
Center Director
V. A. C. Henmon Professor
Educational Psychology

Stephen J. Knezevich
Professor
Educational Administration

Joel R. Levin
Associate Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

Wayne Otto
Professor
Curriculum and Instruction

Thomas A. Romberg
Associate Professor
Curriculum and Instruction

Peter A. Schreiber
Assistant Professor
English

Richard L. Venezky
Associate Professor
Computer Science

Alan M. Voelker
Assistant Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Communication Arts